NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

DEVELOPMENT OF A HUMAN PERFORMANCE MODEL AS A BASELINE FOR AUTOMATIC CHANGE DETECTION SOFTWARE CAPABILITIES IN MINE WARFARE

by

Jason S. Barrett

September 2008

Thesis Advisor:	Lyn R. Whitaker
Second Reader:	Richard D Williams

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2008	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Development of a Human Performance Model as a Baseline for Automatic Change Detection Software Capabilities in Mine Warfare			**5. FUNDING NUMBERS**
6. AUTHOR(S) Jason Barrett			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

This study focused on the development of a human performance model as a baseline performance capability for automatic change detection software for use in mine warfare. Through a series of survey images, operator performance was observed under a variety of sonar image conditions, including increasing clutter levels and changes in image altitude and orientation. While a rough model was developed utilizing only the physical attributes of the images, to obtain a close fit between the model and actual observations, the variability of personal proficiency was included in the final model. The inclusion of this parameter greatly improved model accuracy and highlights the need to better understand differences between operator performances in mine warfare planning.

14. SUBJECT TERMS Mine Warfare, Change Detection, Clutter Density			**15. NUMBER OF PAGES** 51
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

DEVELOPMENT OF A HUMAN PERFORMANCE MODEL AS A BASELINE FOR AUTOMATIC CHANGE DETECTION SOFTWARE CAPABILITIES IN MINE WARFARE

Jason S. Barrett
Lieutenant, United States Navy
B.S., Georgia Institute of Technology, 2001

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 2008

Author: Jason S. Barrett

Approved by: Lyn R. Whitaker
Thesis Advisor

Richard D. Williams
Second Reader

James Eagle
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This study focused on the development of a human performance model as a baseline performance capability for automatic change detection software for use in mine warfare. Through a series of survey images, operator performance was observed under a variety of sonar image conditions, including increasing clutter levels and changes in image altitude and orientation. While a rough model was developed utilizing only the physical attributes of the images, to obtain a close fit between the model and actual observations, the variability of personal proficiency was included in the final model. The inclusion of this parameter greatly improved model accuracy and highlights the need to better understand differences between operator performances in mine warfare planning.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

This study developed a human performance model to determine the minimum performance capability of automatic change detection software for use in mine warfare applications. While the adoption of these software programs is within the foreseeable future, to date, there has been no work in determining what level of performance such a program would be required to meet in order to surpass the capability of human operators performing change detection.

To determine how an operator from the general population would perform when conducting change detection analysis, a survey consisting of ten different side scan sonar images was created. Each image consisted of a "historical" image and a "changed" image which contained two additional objects which the survey participant was asked to identify. In order to establish the effect of environmental and operational factors such as bottom clutter density, changes in sonar height above bottom, and track orientation, and object size, each image was unique with regards to each of these.

While a basic model using only the previously mentioned factors was obtained using S-Plus, in order to create a model which better matched the actual performance observed in the surveys, the additional factors of participant identification and order of performance were also added. These additional factors greatly improved the prediction of performance as compared to the model without the personal factors.

The importance of recognizing the impact of individual ability in change detection applications is key to the development of any standard of performance. Often times in developing performance estimates for mine warfare, only system and environmental parameters are considered. This study demonstrates that while these factors are important, the variability among individual operators is significant. Further study should be given to determining what particular individual traits, if any, account for a specific level of performance in change detection analysis.

THIS PAGE INTENTIONALLY LEFT BLANK

I. THE ROLE OF CHANGE DETECTION IN MINE WARFARE AND HOMELAND SECURITY

Following the attacks of September 11, 2001, the United States undertook an initiative to identify and protect sites and systems that could make inviting targets for terrorist organizations. Almost immediately, symbolic places such as national monuments, and locations where a large a number of casualties could be inflicted, such as professional sporting events, took additional measures to prevent, or at least mitigate terrorist attacks. Unfortunately, other, less dramatic sites, such as transportation lines and public utilities, which were also identified as possible terrorist targets, did not receive the same level of additional security. Central to the daily conduct of business and trade, the destruction of vital infrastructure would have enormous implications on the United States economy as well as cause the American people to question their government's ability to protect them.

Among the top infrastructure and economical concerns was the ability of the United States to protect its port facilities and waterways from attack. With 361 public ports and thousands of miles of navigable waterways spread throughout the country, the task of protecting all of them proved daunting. With a combined economic impact reaching into the trillions, the closure of these facilities and routes would have a devastating effect on the national economy. In 2002, the West Coast dock workers strike closed 29 Pacific ports, costing the American economy nearly $2 billion a day[1]. The effects of this closure showed the ripple effect that a port closure could have across the economy. While the strike made a significant impact on the economy, its effect was somewhat reduced by the fact that the strike had been anticipated. Businesses had been able to find alternate transportation routes, reduce shipments, and create stockpiles prior to the closure of ports. These measures would not be able to be performed prior to a surprise attack, creating even more dramatic economic losses if a port were to be closed.

[1] Grace V. Jean, Improvised Explosive Devices: Could they Threaten U.S. Ports? *National Defense Magazine*, www.nationaldefensemagazine.org/issues/2008/January/Improvised.htm, 02 February 08.

One of the simplest methods by which to attack ports or waterways would be to deploy underwater explosives, either in the form of a traditional military-style sea mine or a "homemade" improvised explosive device (IED). With dozens of countries manufacturing new mines every year for sale and a world inventory of sea mines in the hundreds of thousands, it would be relatively easy for a terrorist group to purchase a mine on the black market, be given mines by a sympathetic state, or simply steal them. Another option for an underwater explosive would be an IED. Capable of being constructed from common items such as fertilizer and fuel, a terrorist could construct an explosive inside the United States and then deploy it in a port or waterway. It is known that numerous terror organizations have extensive explosives training programs and the feasibility of an attack using such an explosive was demonstrated in the Oklahoma City bombing of 1996. An attack of this nature was also demonstrated in a maritime setting in April 2004 when officials in Louisiana recovered and destroyed a garbage bag from Lake Ponchartrain containing several pounds of explosives set to explode with a timer. It is believed that this device had been delivered from a vehicle passing over the lake on a bridge. Had the same device been placed in Los Angeles or New York harbor and exploded, the resulting effort to ensure no other explosives were present may have closed either port for days and cost billions of dollars.

While an actual explosion or the discovery of an explosive device would be an effective means of disrupting maritime commerce, it is possible to achieve the same results with only the threat of an underwater explosive. In January 1980, the so-called "Patriotic SCUBA Diver" crisis closed the Sacramento River for four days while U.S. Navy assets surveyed the river. With only the claim that a mine had been placed in the river, the perpetrators had closed a major West Coast waterway and cost hundreds of thousands of dollars in shipping delays alone[2]

From all indications, the impact of a mining incident in a U.S. port or waterway would be disastrous. With more than 90% of all U.S. trade passing through the nation's ports each year, the closure of any of these facilities would have an enormous effect

[2] Scott C. Truver, Underwater IEDs…The Threat is Real!, 30 October 2007, Presentation to the ASNE Flagship Seminar, Washington, D.C.

throughout the economy. The closure of any single "key" port, such as Los Angeles, New York, or Houston would create untold economic losses through the loss of oil imports alone. While the need to protect these critical assets has been identified, the method through which to achieve this goal remains undecided.

To protect ports and waterways from terrorist mines, the role of the military, specifically the Navy, in homeland security missions has been closely studied. As the only governmental agency with current underwater explosive clearance assets, many officials feel that the Navy is better suited to undertake the protection of domestic ports from this threat than other law enforcement agencies such as the Coast Guard. With a fleet of vessels and aircraft dedicated to mine clearance operations, as well as a number of special operations units trained in such methods, it would make sense to employ these assets to protect vital maritime economic assets. While this idea is initially optimistic, the limitations of such a plan quickly become apparent. With a current inventory of fourteen minesweepers, some of which remain overseas at all times in support of deployed forces, it is impossible to place one in every port. These vessels also have slow transit speeds, which would greatly hinder their movement between ports. These limited speeds could mean days before a minesweeper would reach a port to begin clearance operations, which themselves could take days depending on the ports size and bottom condition. Bottom conditions in a port are particularly important to mine clearance operations. In order to ensure the highest probability of successful clearance, any object that has the appearance of possibly being a mine must be investigated. Decades of accumulated objects such as steel drums, refrigerators, automobiles and other items could add days if not weeks to clearance operations. Every day spent either transiting to, or clearing a port, equates to billions of dollars in lost business as well as the cost in declining public morale.

An alternative to response-based mine clearance operations, is change detection. In this process, routine surveys of a port are made, typically using side scan sonar, and then compared to one another to determine if there has been any change to the objects on bottom. Following a mining threat or incident, a new survey would be conducted and the results compared to those from the most recent survey to identify any new objects that

may be present. These new objects would then be classified as either "mine-like", meaning that they could possibly be mines, or "non mine-like", indicating that they are considered to pose no threat. This method reduces the time required for clearance operations by eliminating the need to investigate every mine-like object such as drums or appliances that may have been previously present. Instead, operators focus on objects that have arisen since the introduction of a threat. Even with reduced time requirements, change detection is a time consuming endeavor. Some estimates conclude that the initial survey of the nation's twenty busiest ports would take three years and cost $14 million[3].

Although there is currently no method in place to perform change detection analysis in U.S. ports, there are initiatives underway to both assess the feasibility of such operations and to develop the required capability to carry these operations out. Recommendations have been made to use both active duty and reserve Navy personnel to perform these surveys, as well as the possibility of contracting such tasks to commercial companies. Utilizing small underwater vehicles equipped with side scan sonar, a limited number of either military or contracted personnel could rapidly survey a port area in the event of a mining incident, rather than waiting for mine clearance ships to arrive. The size of these vehicles would also permit rapid travel between ports by air should a mine threat appear in a port without its own dedicated survey team. Once on scene, operators could deploy their vehicles and commence surveying in the new port.

Even though the speed in which a survey can be performed by underwater vehicles is a dramatic improvement over traditional mine clearance assets, the process of comparing each survey to its predecessor is extremely time consuming. These comparisons are typically done by operators who visually compare images. In an environment with numerous bottom objects, this task can be daunting. Numerous factors can contribute to the difficulty in performing change detection. Factors that increase the time required for an operator to correctly identify changes include: the number of objects: the relative orientation of the images to each other, and changes in both sonar system and environmental conditions, such as the height above bottom the image was recorded at and

[3] Grace V. Jean, Improvised Explosive Devices: Could they Threaten U.S. Ports? *National Defense Magazine*, www.nationaldefensemagazine.org/issues/2008/January/Improvised.htm, 02 February 08.

object movement. As a result of this, projects are underway to develop computer software which can perform change detection analysis. One method being investigated, allows new objects to be identified in a real-time fashion as the survey is being performed. A historical survey is loaded into the sonar control software and the historical image is compared to the current survey as it is being performed. The algorithm compares the objects at each geospatial coordinate (obtained by GPS) to the objects at the same GPS position in the historical image. The second method involves the same use of GPS positions, but the comparison between the current survey image and historical images is done at the completion of the new survey. In order to match an object to an object in a previous survey, the position as recorded by the software's navigational component must be 100% repeatable[4]. Current navigation systems, while very close, hold some intrinsic error in their positions, making point-for-point comparisons impossible. Although this issue could be overcome with object shape comparison, such a process could be performed only on the clearest of sea floors, minus any similar shaped objects.

While it is probable that in the future automated software programs will be developed to rapidly perform change detection, the best option currently is to utilize human operators. This thesis will research the capabilities of these operators in order to determine to what standard an automated program must perform in order to exceed our current capabilities.

The preparation and conduct of this study, along with the results, conclusions, and recommendations of the study will be discussed as follows: Chapter II will describe the Development of Change Detection Scenarios, Chapter III will discuss the results of the study survey and develop a human performance model, and Chapter IV will present conclusions from the study, along with recommendations and ideas for future work in the area of human/ACD software comparison.

[4] Gary Kozak, Side Scan Sonar Target Comparative Techniques for Port Security and MCM Q-Route Requirements, L-3 Communications Klein Associates, Inc, 2006.

THIS PAGE INTENTIONALLY LEFT BLANK

II. THE DEVELOPMENT OF CHANGE DETECTION SCENARIOS

A. CHANGE DETECTION METHODOLOGY

The performance of visual change detection is most commonly performed by one of two methods, either a side by side comparison of images, or through the utilization of transparent overlays. Both methods have advantages and disadvantages according to the situation, and the choice of which methodology is often left to the operator. The following paragraphs highlight some of the differences between the two methods.

1. Side by Side Comparison

When performing change detection analysis through side by side image comparison, an operator places the most recent historical sonar image next to the recently acquired image and differences between the two are noted. While differences may exist where an object that was present in the historical image is no longer present in the current image, the operator is mainly looking for objects that are present in the current image that are not in the historical. The presence of a new object could indicate that an underwater explosive has been introduced. The decision as to whether or not a new object could be a mine is based on the object's size, shape, and other factors corresponding to the characteristics of the anticipated threat. Such an object is known as a mine-like object (MILO). An advantage to this methodology is that both images can be displayed on a single computer monitor and compared, avoiding any distortions or decreasing the level of detail that may result from printing the image. This also allows for one continuous image to be viewed through the use of scrolling rather than viewing a segmented image that would result from the printing process without a specialized printer. One disadvantage of this method however is the requirement that follow-on surveys be orientated exactly as the historical survey. Any deviation in track direction results in the new and historical images being out of "sequence", that is, they cannot be compared by

simply scrolling through both together. Instead, a point by point comparison of coordinates is required, vastly increasing the time required to complete the analysis.

2. Overlay Comparison

Overlay comparison involves using transparent overlays to compare changes in bottom objects. Typically, a previous survey image is placed overtop the new image and any objects that show through the historical image from the new image are marked and their image is reviewed to determine if they are a MILO. This method has the advantage of being able to overcome differences in survey orientation and object movement. For example, if an initial survey is performed along a north-south axis, the pattern of objects on the bottom will be orientated in a particular pattern. If a follow-on survey is then performed along a different axis, the orientation of objects will appear different. The use of overlays allows the operator to "twist" the images in order to match their orientations for comparison. The effect of object movement can also be mitigated through this process as the new image can be shifted to overlay the original image, assuming that an object can be identified as the same in both images. This is of particular benefit in areas where objects move or "walk" at a known rate. One disadvantage of this method though is the requirement to have available overlay transparencies, printers, and organizers, greatly increasing required space and introducing the possibility for errors due to poor organization that are not present in completely electronic methods.

B. EFFECTS OF SIDE SCAN SONAR EMPLOYMENT IN CHANGE DETECTION

Ideally, every survey of an area would be performed under identical conditions. While environmental factors such as attenuation and currents do affect images obtained from side scan sonar's, the major source of difference in detail level among subsequent surveys is the height above bottom of the sonar during each survey and the speed at which the sonar moves through the water. While these parameters are within the control of the sonar operator, experience level, time constraints, and platform sometimes prevent identical conditions. The following paragraphs give the details of the effect of these employment characteristics.

1. Height Above Bottom

A side scan sonar moving through the water can be likened to an airplane moving through the sky. As anyone who has flown can attest, the higher the plane above the ground, the larger an area can be seen and the smaller individual objects appear. This same effect occurs with side scan sonars. The greater the height above bottom, (also known as altitude), the wider the potential swath of sea floor that will be visible in the recorded image. While altitude is important to determining swath width, frequency and range setting also play important roles. For the purposes of this research however, these factors will be ignored. Differences in swath width between surveys is important because of the impact it has on relative object size. If an object is viewed as part of a 50 m swath, the same object will appear smaller when viewed as part of a 75 m swath. The actual percentage of change in relative size depends on the angle of the sound beam striking the bottom at the edge of the swath. This change in observed size plays an important role in determining the presence of new objects in surveys. If a survey is initially performed at one altitude and subsequent surveys are performed at a greater altitude, bottom objects may appear significantly smaller, possibly to the point that new objects are so small as to avoid detection. Conversely, if an initial survey is performed at one altitude and then subsequent surveys are performed at a lower altitude, objects that were previously undetected may now appear large enough to be detected and be reported as a change.

2. Side Scan Sonar Speed Through the Water

The resolution of a side scan sonar image is greatly dependent upon the speed at which the sonar moves through the water relative to the swath width. The larger number of returns from an object, the better the image obtained. The period of time in which the sonar is performing as either a transmitter or receiver must be carefully matched to the speed of the sonar through the water. Sufficient time must be allowed for the sound energy to travel to the bottom and then be reflected back. If the sonar is moving too quickly through the water, some return pulses will be missed as the sonar body will have moved out of the line of return. This situation is demonstrated in Figures 1 and 2.

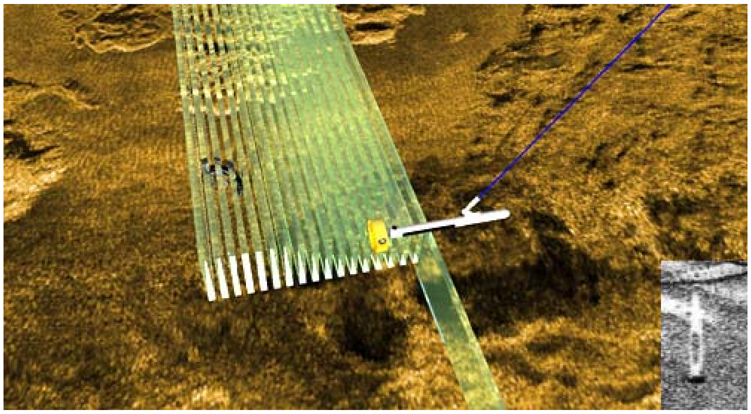

Figure 1. Side scan sonar traveling at appropriate speed to obtain complete coverage.

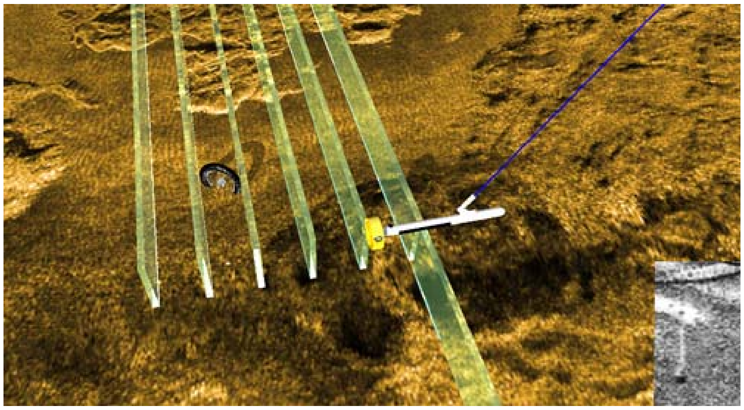

Figure 2. Side scan sonar traveling too fast to obtain complete coverage.

C. DEVELOPMENT OF SIDE SCAN SONAR IMAGES FOR CHANGE DETECTION ANALYSIS

The purpose of this research is to determine a baseline performance capability for automated change detection software in regards to the ability to detect new objects placed in a survey area under various clutter conditions. The Navy currently identifies three different clutter categories based on the density of non-mine bottom objects (NOMBO) per square nautical mile. These categories are defined as follows:

NOMBOS/nm^2	Clutter Category
< 15	1
15 – 40	2
> 40	3

Table 1. U.S. Navy Clutter Categories

It is important to note that the NOMBO density is not related to the MILO density of an area, as the determination as to whether a particular object is "mine-like" is left up to the individual operator. NOMBO density is simply a measure of the number of objects on the bottom that will produce a sonar return.

While there are numerous factors that can impact an operator's ability to discern changes in bottom surveys, this project will limit its scope to the effects of increasing clutter density, object orientation, and scaling as a result of sonar altitude. To test the role of these factors in operator performance, survey images were created utilizing available side scan images, and then altering the image to produce a change. All images were created using Microsoft Paint.

1. Base Image

In order to establish a standard bottom on which to test change detection performance, a "clean" bottom, clear of any visible objects, was created. A sample side scan sonar image, provided by Klein Associates, was selected based on its clarity and bottom composition. The selected image was taken over a hard sand bottom which provides the best surface to avoid object burial and excessive returns. The side scan sonar used was the Klein 5000 system, using a 75 m range, a tow speed of 7.5 kts., and pulse frequency of 455 kHz. Sonar altitude was approximately 10 m. In order to "clean" the image of objects, the image was opened in Microsoft Paint and a small section of sand next to any object was copied and placed over the object. This method allowed for maintaining the natural shading and contours of the bottom. A section of the image

equating to a 150 m by 90 m area was used to allow for the image being printed on a single page. The original and "cleaned" images are shown below.

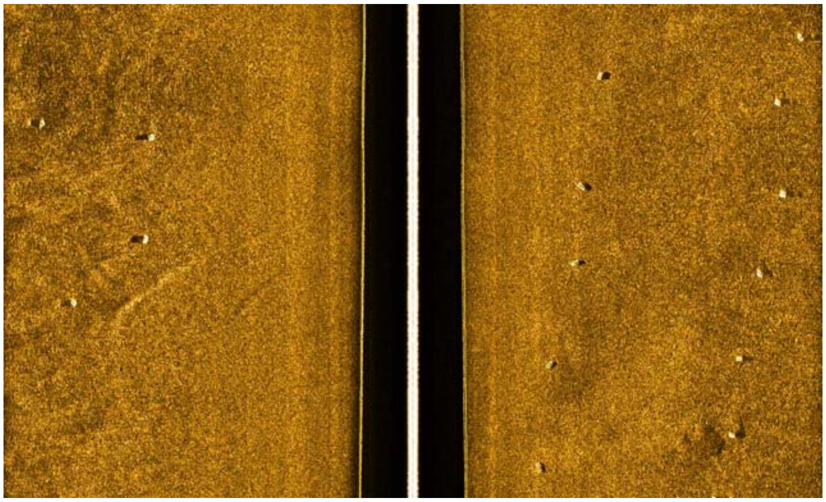

Figure 3. Original side scan sonar image.

Note the numerous objects on both sides of the sonar track in Figure 3.

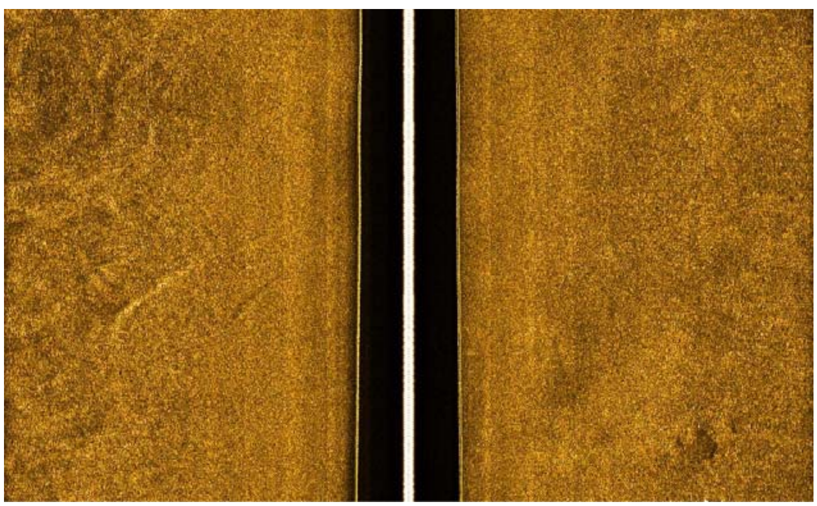

Figure 4. "Cleaned" side scan sonar image.

Note that clutter objects have been removed, while maintaining natural features such as sand ridges.

2. Clutter Images

In order to determine operator performance under differing levels of clutter, it was necessary to devise a method for distributing objects throughout the base image. While Navy clutter categories are based on square nautical miles, that scale would be insufficient to show any difference in the limited bottom area used in this research. A Clutter Category of 3 (40 NOMBO/nm^2) would indicate an average of 1.16 * 10^{-5} objects per square meter. With a total area of 13,500 m^2 in our image, this would equate to 1.35 objects throughout, hardly a basis for comparison. For the purposes of this thesis, we will use clutter densities of NOMBO/100 m^2. Six categories of clutter were then defined as follows:

Clutter Category	NOMBO/100m^2	Total Number of Objects
1	0.07	1
2	0.5	7
3	1	14
4	1.5	20
5	2	27
6	2.5	34

Table 2. Image Clutter Density Categories

These categories were selected in order to provide a noticeable difference in clutter densities, while keeping the total number of objects at a level that allowed each object to reside in its own unique point. Side scan sonar images of three different objects — a round crab pot, a square crab pot, and a cement block — each taken from images recorded with the same parameters as the base image, were then inserted at random points into the base image. These points were selected using Microsoft Excel's Random Generator. Since the base image was opened in Microsoft Paint, each point had a unique coordinate comprised of its x and y pixel values. X values ranged between 0 and 1132 and Y values ranged between 0 and 688. The type of object to insert was also chosen at

random by Microsoft Excel, using values of 1, 2, and 3 respectively for each type of object. An example image with an inserted object is shown in Figure 5.

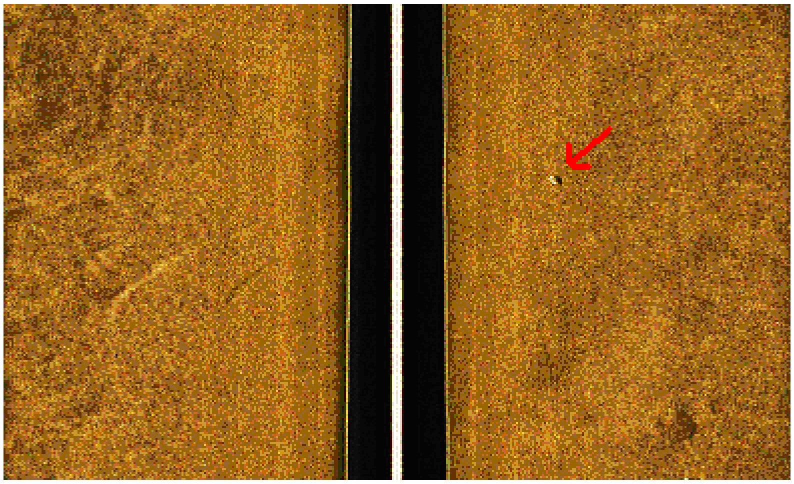

Figure 5. Simulated side scan sonar image showing object type 2 (square lobster pot) placed at coordinates (793, 251).

Images for each clutter level were created in the same manner. Copies of each image had two new objects inserted in the same manner to provide a "change" which each operator then attempted to locate. All images were produced on both paper and transparencies in order to allow research participants to perform both side by side and overlay comparison techniques. Participants were asked to indicate the presence of any new objects in each image by circling them.

3. Change of Orientation Images

In order to determine the effect of altering the track of a side scan over the same area and the corresponding change to the orientation of the survey image, a series of images were altered to achieve this effect. Using clutter categories of 2 and 3, two initial images were created in the same manner as before. These image orientations were then rotated by 90 degrees by removing the sonar track, and replacing this area of the image with a "clean" sand background. The sonar track was then returned to the image, perpendicular to its initial direction. Particular care was given to ensure that the two additional objects were then added to the new image following the same method as before. An example of the resulting image is shown in Figure 6.

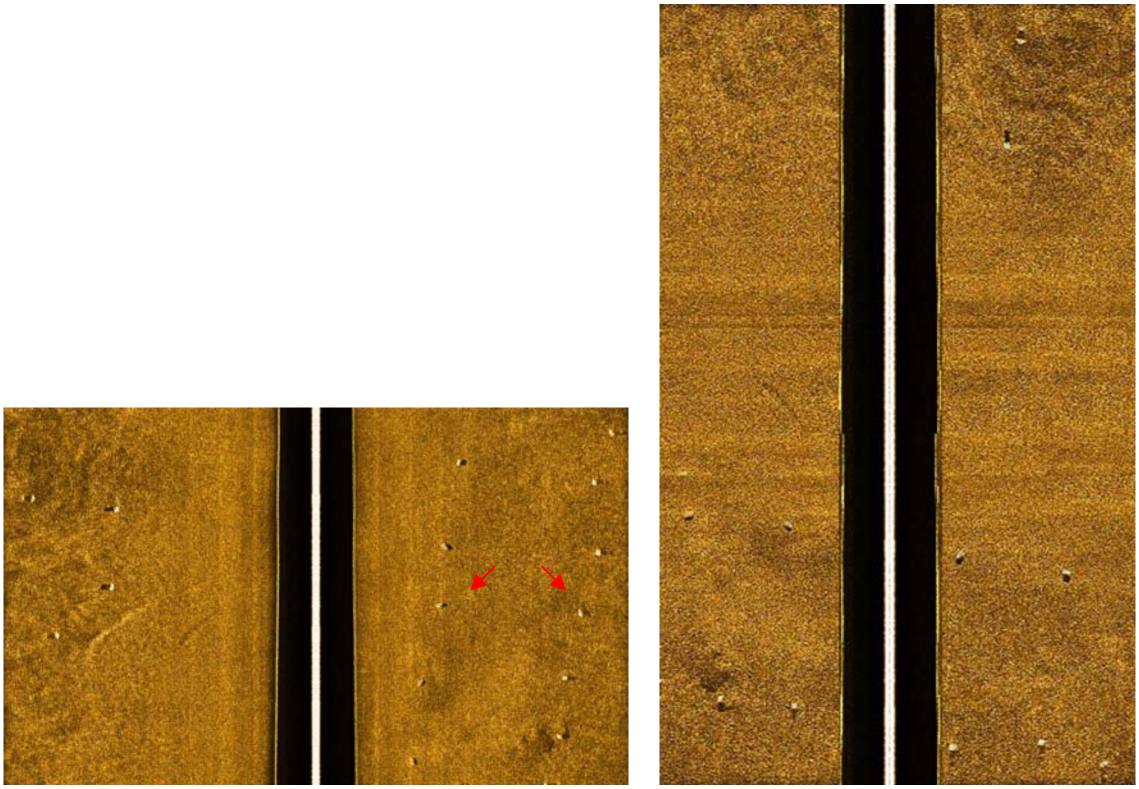

Figure 6. Original and orientation altered survey image.

Note that the new track covers some of the original bottom objects.

4. Change of Altitude Images

In order to determine the effects of increasing sonar altitude and the resulting reduction in visual object size, a set of images corresponding to a 25% reduction in scale were created. Two images, with clutter categories 2 and three were created in the same manner as previously discussed. Copies of these images were then reduced to 75% of their original size, and “clean sand” was added to the images to create the appearance of increased area. Two new objects, scaled to match the increased area, were then inserted to provide “change” for analysis. An example of these images is provided in Figure 7.

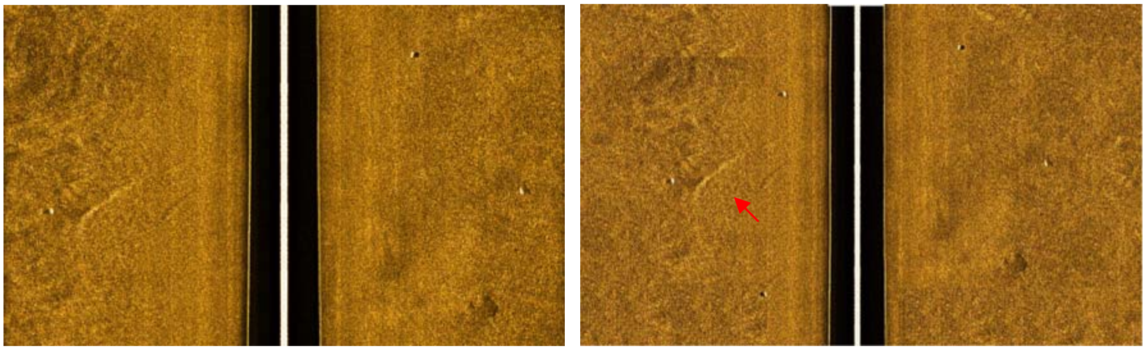

Figure 7. Original and altitude altered survey.

Note the sand ridge in the bottom right corner of the original image has moved slightly to the left and up in the altered image.

III. ANALYSIS OF DATA

A. SURVEY RESULTS

A series of 50 surveys were distributed to Navy Postgraduate School (NPS) students and the public, of which 31 were returned. Each survey was identical, except for the random assignment of each image to be printed on either paper or a transparency. This allowed users to perform change detection analysis utilizing either side-by-side comparison on paper images, or overlay comparisons on the transparencies. Before distribution, the survey format and images was approved by the NPS Institutional Review Board. Each participant returned a signed consent form along with their survey, however, no information connecting an individual to performance was recorded. Each survey image was reviewed against a master image to determine if the participant had correctly identified the new objects found in the image, assigning a value of one to those objects found, and a value of zero if the object was missed. The conditions of each image such as clutter level, object type, altitude and orientation change, and method of analysis was recorded along with whether or not the desired object was found. Using this method, a total of 620 data points were created. A summary of these data was then created, as shown in Table 3.

Clutter Category	Percentage Found	Object Type	Percentage Found	Detection Method	Percentage Found
1	0.58	A	0.37	Side-by-Side	0.61
2	0.63	B	0.30	Overlay	0.53
3	0.89	C	0.87		
4	0.61				
5	0.87				
6	0.52				

Table 3. Summary of Change Detection Survey Results

For further analysis, the survey results were then loaded into the statistical software package S-Plus in the format shown in Table 4.

Cltr	Object	Fnd	ID	ORDER	Alt	Orn	Over
1	A	0	A	1	0	0	0
1	A	1	B	1	0	0	1

Table 4. Example of Data Format Loaded into S-Plus

The values of Cltr (Clutter Category) correspond to those provided in Table 2. Object (Object Type) denotes which of the three objects the participant is expected to locate. Fnd (Object Found) is a binary variable with either a value of 1 or zero, depending on whether the object was found or not. ID indicates which participant attempted to locate the object, although no individual is connected to any particular identification code. ORDER is the presumed order in which the participant viewed the image within the series. While participants were not directed to analysis the survey images in any particular order, nor were they asked to record in what order they analyzed the images, it is assumed that all participants performed the required image analysis following the sequential numbering of the images themselves and in the order each image was described in the survey instructions. Alt (Altitude Change) is a binary variable, assigned either 1 or 0 indicating whether or not the image corresponded to a change in sonar altitude. Orn (Orientation Change) is also a binary variable, indicating an orientation change in the image. Over (Overlay) indicates whether or not the image appeared on a transparency and was therefore analyzed using the overlay technique.

B. MODEL DEVELOPMENT

To examine the data, the sample prediction of detections (p-hat) for each object type by clutter category was calculated and plotted against clutter categories in Microsoft Excel to determine the existence of effects from physical survey variables. P-hat values which remained constant suggest that the variable under consideration may not impact survey performance by itself. P-hat values that fluctuate suggest that the variable under consideration does impact survey performance. Figures 8 through 10 show these relationships:

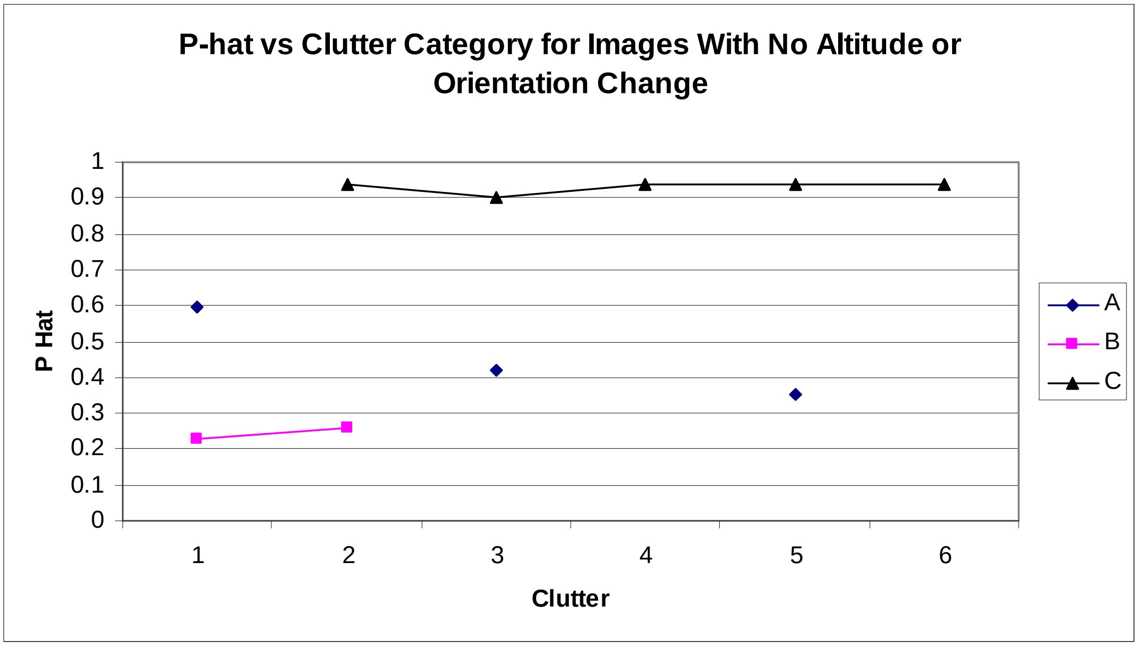

Figure 8. P-hat vs Clutter Category for Images With No Altitude or Orientation Change

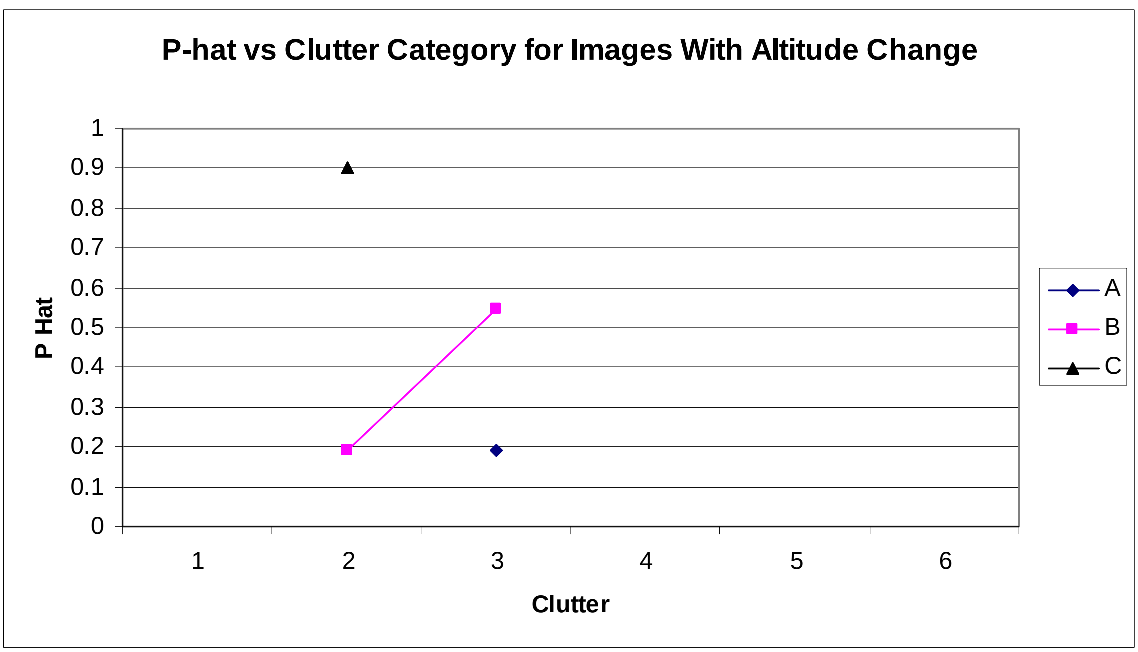

Figure 9. P-hat vs Clutter Category for Images With Altitude Change

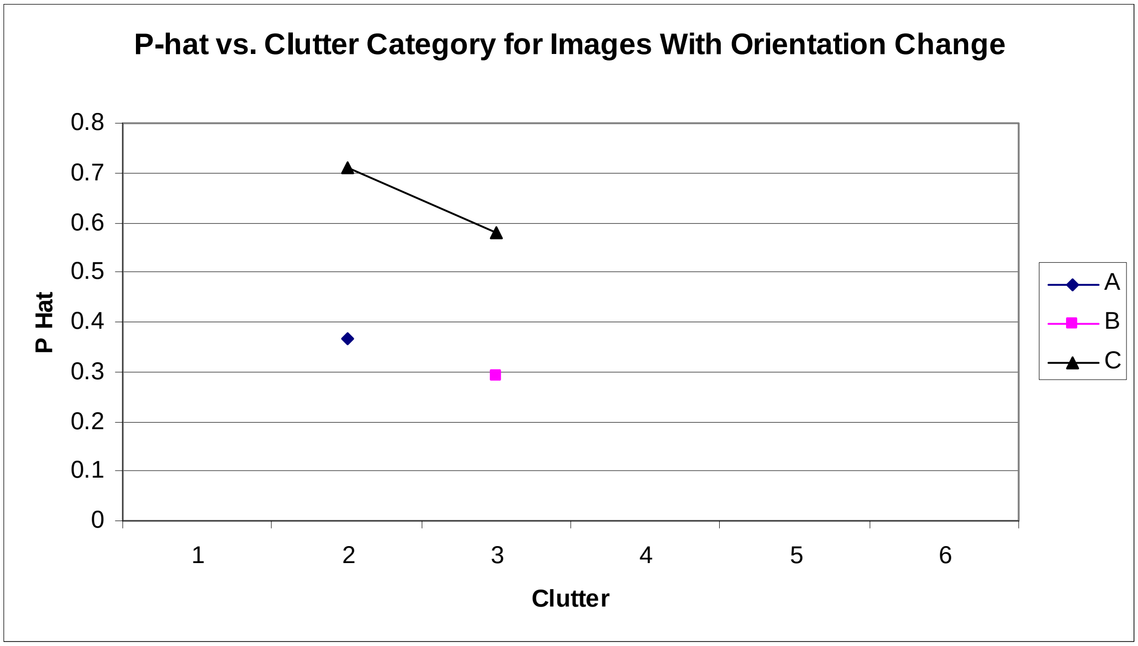

Figure 10. P-hat vs Clutter Category for Images With Orientation Change

As can be seen in Figure 8, increasing clutter density has a negative effect on the probability of detection for the smallest object (Object A), but little effect on the largest object (Object C). However, increasing clutter density appears to have a slight positive effect on the probability of detection for the mid sized object (Object B). Figure 9 shows a dramatic increase in the probability of detection of Object B when subjected to an altitude change while Figure 10 shows a sizable drop in the probability of detection for Object C when viewed with an orientation change. These trends suggest the possibility of an interaction between object type and the variables Clutter, Altitude, and Orientation.

It is important to note that in Figures 8-10, line segments are not smooth due to the integer values of clutter categories. Also, not all image traits were present over all clutter categories for all objects, creating some single point data series.

This initial analysis suggested that the variables Object, Clutter, Altitude, and Orientation should be considered in the model. Once contributing physical survey traits were determined, an initial model was created in S-Plus to encompass the possible effects of both independent variables and their interactions. The response variable, Fnd, is a Bernouli variable taking values of 0 or 1 as previously described. As such, the model

was assumed to be a logistic regression model.[5] In particular, let n be the number of observations, then let Y_i with i = 1..n represent the binary variable Fnd. The logistics regression model assumes that $Y_i \sim$ Bernouli(P_i) where $P_i = P(Y_i = 1)$ for i = 1...n, where $Y_1..Y_n$ are independent and that the distribution of the response variables is linked to the explanatory variables through log-odds:

$$\log(P_i/1-P_i) = \beta_0 + \beta_1 xi_1 + \ldots \beta_k xi_k$$

where $\beta_0 \ldots \beta_k$ represent the coefficients corresponding to the explanatory variables $x_{i1} \ldots x_{ik}$ for i = 1…n. In this analysis, some of the explanatory variables are binary (such as Alt and Orn); some are numeric such as those for Clutter and Order; and the categorical variables with l levels (ID with 31 levels and Object with three levels) are represented y l-1 categorical variables

From this initial model, the S-Plus automated stepwise regression function, stepAIC[6] developed a consolidated model to predict the success or failure of detecting a new object. This function performs a stepwise regression using a backward elimination, removing variables and interactions that are found to not significantly contribute to the model prediction value.[5] The stepAIC function recommended the following prediction model which included the variables Clutter, Object, Order, Altitude, Orientation, and the interaction between them. S-Plus also provided coefficients for the prediction model as shown in Table 5.

[5] Jay L. Devore, Probability and Statistics for Engineering and the Sciences. Belmont: Thomson, Brooks, Cole, 2004.

[6] W.N. Venables, B.D. Ripley, Modern Applied Statistics With S. New York: Springer, 2002.

Term	Coefficient	Standard Error
Intercept	-1.189939	0.721035
ObjectB	-0.09552016	1.276822
ObjectC	2.436686	0.910995
Cltr	3.199586	1.151499
ORDER	-1.718773	0.571076
Alt	4.424374	6.285507
Orn	25.06601	8.202927
ObjectB:Cltr	0.4416618	0.480964
ObjectC:Cltr	0.323061	0.206186
ObjectB:Alt	1.63891	0.745733
ObjectC:Alt	1.698297	1.158843
ORDER:Alt	0.768086	0.369562
Cltr:Orn	-0.8528581	0.467001

Table 5. S-Plus stepAIC Recommended Prediction Model Coefficients

Note that the coefficients for ObjectB corresponds to a binary explanation variable which takes a value of 1 if Object B was present and 0 otherwise. Similarly, for ObjectC, the coefficient corresponds to the binary variable which takes a value of 1 if Object C is present and 0 otherwise. Values of 0 for both Objects B and C indicate that Object A is present. When multiplied by the appropriate variable values, these estimated coefficients yield estimates of the log odds which can then be translated into a predicted probability of detection for a set of given conditions. For example, if Object B is present, in a Clutter Category of 4, with an apparent altitude change between images, and the image is the fifth one viewed by the participant, the model equation would be:

$$\log(\text{P-hat}/(1-(\text{P-hat}))) = -1.189939+-0.09552016+3.199586(4)+-1.718773(5)+4.424374 +0.4416618(4)+1.63891+0.768086(5)$$

$$\log(\text{P-hat}/1-(\text{P-hat}))=14.59$$

$$\text{P-hat}=0.34$$

Indicating that under these conditions, the predicted probability of detection would be 0.34. A plot of the predicted probability of detection values, and the probability of detection values from the survey is given in Figure 11.

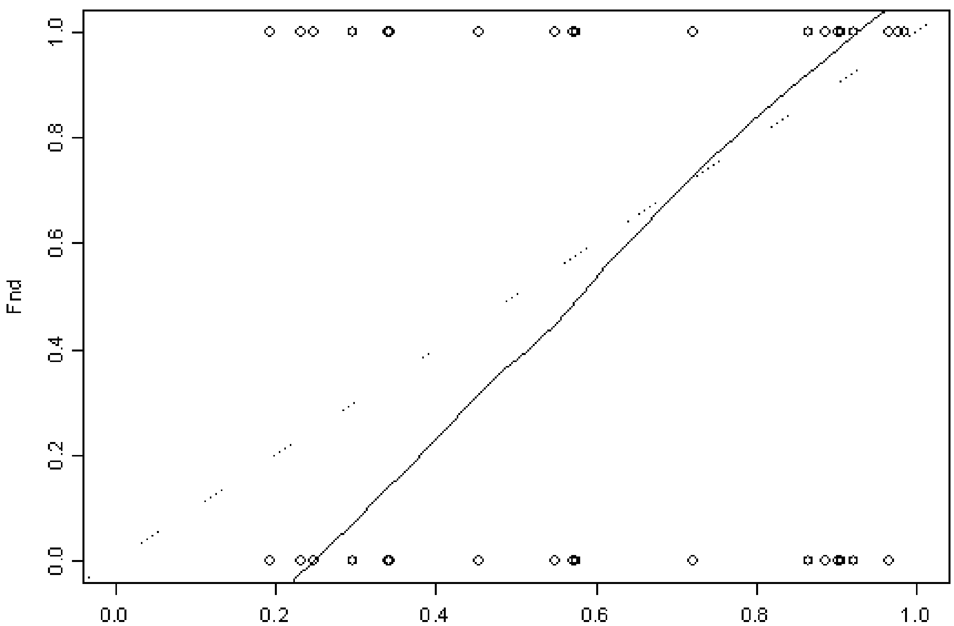

Figure 11. Plot of the Response Variable vs. the Estimated Probability of Detection for Model 1

In Figure 11, the dashed line is the identity function and serves as a frame of reference. The solid line is a smoothed version of the sample proportion of detections vs. the estimated probability of detection from the actual survey data. Since Fnd is a binary term in the data set, the sample probability of detection for the data is an average of all data points with the same conditions.

As can be seen from Figure 11, the predicted and actual values follow the same trend, but are noticeably different across much of the range of predicted values. A number of estimated coefficients are also suspicious in this model, namely the estimated coefficient for clutter. Generally, it is understood that an increased clutter level should have a negative impact on the ability of an operator to identify a new object in a sonar image. However, in this model, clutter has a positive coefficient, indicating that increasing clutter levels actually make the identification of new objects easier when considered with all other variables. In an attempt to more accurately match the prediction values of the model with the actual survey observations, the effects of personnel were included and tested. This model retains the original terms from the first model, but also includes the effects of ID and the effects of the interaction between Object and ID.

Executing the model in S-Plus, 130 unique coefficients were calculated to include the interactions between every survey participant and each object type. Table 6 shows the new coefficient values for this model, less the coefficients for ID and ID interactions.

Term	Coefficient	Standard Error
Intercept	-8.58168	74.66609
ObjectB	4.807992	74.67846
ObjectC	13.58109	74.69453
Cltr	-1.1684	0.767215
ORDER	-0.01696	0.111884
ObjectBCltr	3.163356	0.606853
ObjectCCltr	0.685227	0.330696
ObjectBAlt	0.808273	0.864277
ObjectCAlt	1.756599	0.94787
ORDER:Alt	-0.11654	0.080048
Cltr:Orn	-0.88933	0.63293

Table 6. S-Plus Determined Coefficient for Second Model

The likelihood ratio test of the first model (the null hypothesis) versus the second model (the alternative hypothesis) yields a test statistic with a value of 242.9, which is the difference in the residual deviances of the two models. Under the null hypothesis, the likelihood ratio test statistic has an approximate Chi-Squared distribution with 120 degrees of freedom (the difference between the number of coefficients in the two models)[7]. This gives a P-value of less than 0.1%, indicating that there is strong evidence to reject the null hypothesis in favor of the model which includes ID and its interactions. Figure 12 shows the improved performance of the second model in estimating detection.

[7] Jay L. Devore, Probability and Statistics for Engineering and the Sciences. Belmont: Thomson, Brooks, Cole, 2004.

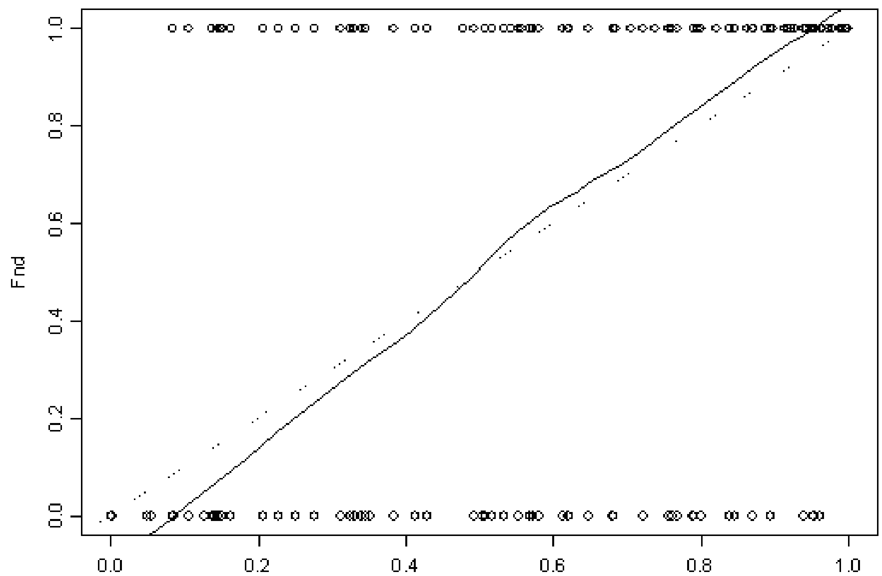

Figure 12. Plot of the Response Variable vs. the Estimated Probability of Detection for Model 2

A second measure of how well the model fits the actual data is the Misclassification Rate. This value relays the percentage of a models calculated probability of detection that would indicate a different result than the observed success or failure of detecting an object. For example, if a model indicated that an object should be found (probability of detection > 50%) based on the image parameters, but the survey participant failed to detect the object, that would be a misclassification. With a misclassification rate of 0.1613, the second model's rate is much lower than the first model's misclassification rate of 0.3935. With models that have many parameters such as the second model, there is always the concern that the model is over fit, i.e. it predicts the data used to fit the model very well (too well), but is not useful for predicting new observations. To check whether the second model fit was too good, a cross-validated misclassification rate[8] was computed to be 0.2419, which is close to the observed misclassification rate. A misclassification rate much higher than the observed rate would indicate over fitting, which is not the case for the second model fit.

[8] W.N. Venables, B.D. Ripley, Modern Applied Statistics With S. New York: Springer, 2002.

C. REASONS FOR MODEL DIFFERENCES

By adding the personnel identification term to the model, the inherent difference between individual proficiency in any task was introduced. While often considered to be marginally important in prediction models, the fact that some people perform particular tasks better than others cannot be ignored. Figure 13 shows the predicted probability of detection of Object A for each survey participant across all clutter conditions, absent any other factors. If all or most individuals performed similarly in each clutter category, the plot would show a tight band of predicted probability of detection values. However, it is clear that the trend of predicted probability of detection values varies widely with each person. For example, participant "T" maintains a fairly constant level of predicted detection across all clutter categories. Participant "A"s predicted detection rate drops sharply in clutter conditions three and higher. Participant "H" by contrast is predicted to do poorly in clutter conditions one through three, but dramatically improves in clutter condition four through six.

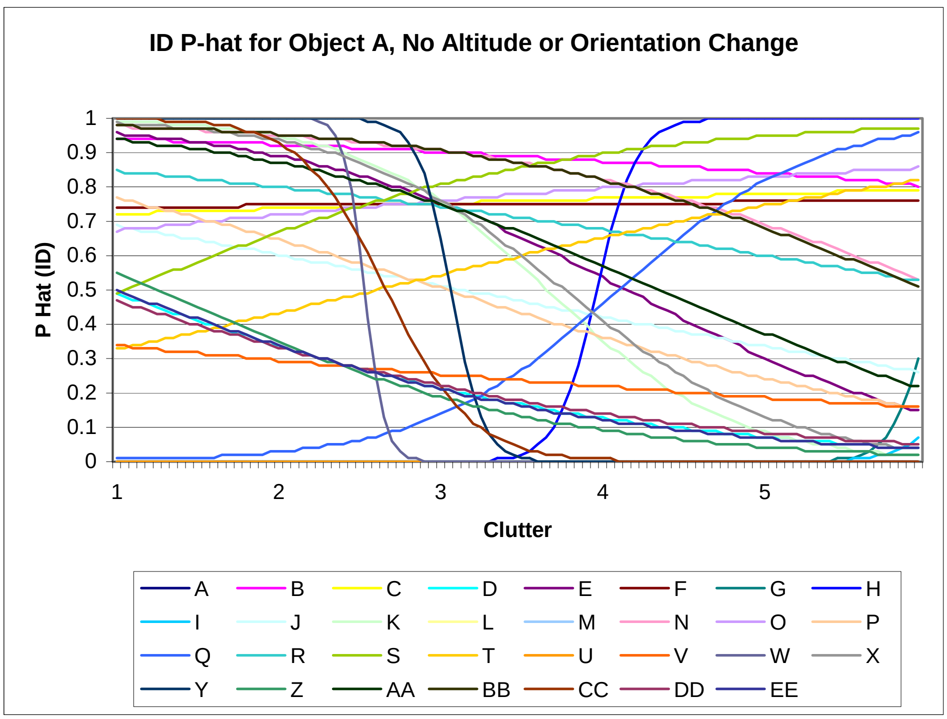

Figure 13. ID P-hat for Object A, No Altitude or Orientation Change

While the ability to identify Object A varied widely, less variability was present among participants in their ability to identify Objects B and C. Indicated by the more compacted nature of the plots for each participant, this phenomenon is most likely due to the increasing size of these objects over Object A. The specific dimensions of each object will be discussed in Chapter IV.

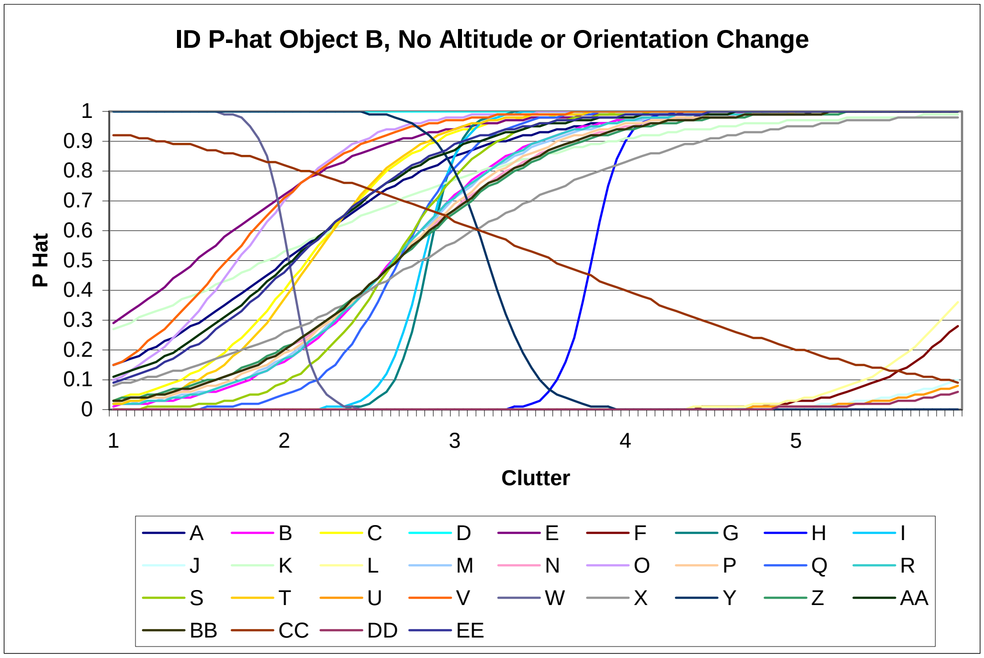

Figure 14. ID P-hat Object B, No Altitude or Orientation Change.

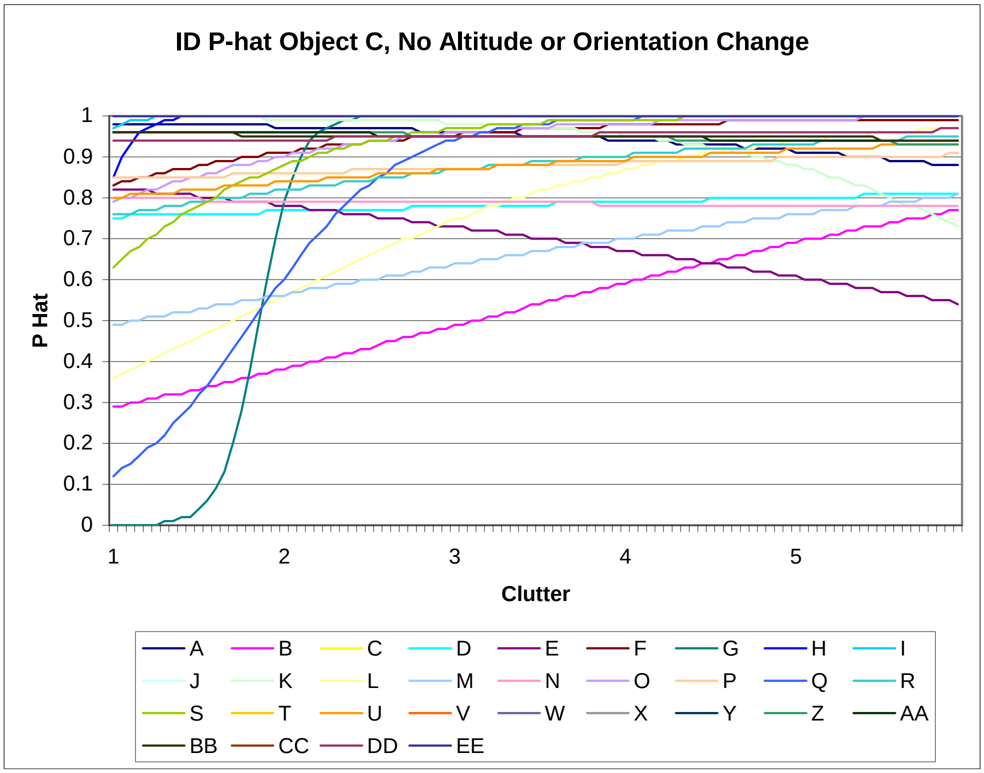

Figure 15. ID P-hat for Object C, No Altitude of Orientation Change.

While it can be seen from the above plots that the performance of the 31 individuals varies widely, this data set is only a small sample of an essentially infinite population of potential operators. In order to better model the effects of different levels of personal proficiency, it would be best to model ID as a random variable. To do this, a sizable population would need to be studied and the effects of each person recorded. S-Plus could then inject this factor, according to the corresponding distribution using the Non-Linear Mixed Effects Model (nlme). This function allows for both fixed value parameters and random variables to be evaluated within the same model.[9].

[9]W.N. Venables, B.D. Ripley, Modern Applied Statistics With S. New York: Springer, 2002.

IV. CONCLUSIONS, RECOMMENDATIONS AND FUTURE WORK

A. CONCLUSIONS

Based on the results of this study, it is possible to model human change detection performance as a baseline for automatic change detection software. These models however must be narrowly tailored to account for anticipated objects, environmental conditions, and the abilities of an operator that could reasonably be expected to perform such an analysis.

One goal of this study was to determine if the method of image analysis, whether it be comparing images side by side or through the use of overlays, produced different probabilities of detection. Based on both regression analyses of the data, and a comparison of detection rates, the method of visual comparison played no discernable role in determining the probability of detection under the survey conditions.

Contrary to expectations, survey participants actually performed better under more demanding conditions. Only when viewed in the presence of all other variables did clutter value appear to have a negative effect on locating new objects. Changes in altitude between images actually appeared to improve performance in the final model. While unexpected, this finding could be explained, at least to some degree, by the principle of underwork and overwork. Researches in the fields of human factors and psychology have observed that the performance level of some people follows a curved path, tasks that are perceived as trivial or unchallenging receive little effort and therefore are performed poorly. The same level of performance is seen when a task is viewed as overly complicated. Subjects feel as though in spite of there best efforts, a task is impossible to complete successfully, and therefore devote little energy to completion[10][11].

[10] Wendelin Schnedler, Task Diffuicuity, Performance Measure Characteristics and the Trade Off Between Insurance and Well-Allocated Effort. www.bristol.ac.uk/cmpo/publications/papers/2006/wp147.pdf. 05 August 2008.

[11] Guido H. E. Gendolla, et al. Self-focus and task difficulty effects on effort-related cardiovascular reactivity, *Psychophysiology,* Vol. 45, 12 February 2008.

Only those tasks which are viewed as sufficiently challenging yet possible to complete receive a sufficient level of effort. The increase in bottom clutter or change in scale due to altitude change could cue survey participants that a particular image is more likely to have a "change" than an image with a lower level of clutter and therefore cause a rise in the level of care and attention given to images with higher clutter values or that have a constant scale. It is impossible to say however, at what clutter level or change in altitude a decline in performance may be seen as either the identification of new objects becomes more difficult, or at least the perception of the task does.

The size of a "new" object inserted into each image greatly impacted the likelihood that it would be detected. The largest object, a 1.1 m squared, 45 cm high square lobster trap was found at a rate more than twice the next smallest object, a one meter long, half meter wide, 30 cm tall cement block. The smallest object used in the survey, a one meter diameter crab pot, only 25 cm in height, was also found at nearly the same rate as the cement block (29% vs 34%). The most notable difference between the appearances of these objects was the large acoustic shadow cast by the lobster trap. Both the crab pot and cement block, with their smaller shadows, proved to be more difficult to detect, confirming the commonly held notion that "proud" objects are easier to find than those flush with the bottom.

One very prominent factor in this model, while often overlooked in mine warfare planning, was individual performance. While most MCM planning guides rely on sonar system performance characteristics to determine a probability of detection in various clutter and bottom conditions, this study revealed that even under identical circumstances, the ability of individuals to detect objects varied widely, in some instances, making the largest contribution to whether or not an object was found. The probability of detection for individuals varied widely, with some always near zero and others always near one. Some individuals demonstrated a linear relationship between performance and clutter, with both positive and negative trends being present. Others demonstrated a more asymptotic relationship, hovering near one or zero over a series of scenarios, and then rapidly moving to the other extreme. The greatest variability between

personal performances was found when searching for the smallest object, the crab pot, and the least variability when searching for the largest object, the lobster pot.

In summary, from this study, we can conclude that it is possible to model human performance to create a baseline performance for automatic change detection software. Key to the development of any model however, is the understanding of the individual performance of likely operators in a particular change detection environment.

B. RECOMMENDATIONS

Based on the results of this study, the following recommendations are made:

- Develop a method to include personal performance in probability of detection calculations. Standardized personnel performance estimates may be based on such measures as years of experience performing change detection, formal training, and the sonar system employed. These estimates should be specific enough to account for the particular environment (i.e. bottom type, burial rates, etc.) in which any operation will take place.
- Develop individual human performance models for each unique environment in which operations may take place. Compare these models to an automated change detection software's performance model for the same environment, realizing that in each situation, a human operator, or an automated system may perform better than the other.
- Consider assigning Navy MCM personnel to monitor a single, or limited number of locations for change detection purposes. Increased experience in a particular environment would likely increase familiarity and result in the more likely detection of new objects. This may require alterations to the Navy's policy of rotating personnel through a number of assignments in varying locations.

C. FUTURE WORK

While this study did find that it is possible to develop a human performance model as a baseline for an automatic change detection software package, more research is required to refine such a model. The following are recommendations for continuing work in this area:

- Determine what personal metrics, such as years experience, age, or formal training impact the ability to identify objects during change detection.
- Investigate any differences in detection rate when survey participants view the same image on both a sonar system display screen and a printed image.
- Attempt to determine the smallest sized object individuals can identify during change detection analysis.
- Study the effects of increasing clutter levels on detection rates using a larger range of clutter values and a variety of different sized and shaped objects.
- Compare both human and change detection software performance over a series of identical scenarios.
- Evaluate theoretical human performance models against actual performance under differing conditions.

LIST OF REFERENCES

Devore, J.L., (2004). Probability and statistics for engineering and the sciences. Belmont: Thomson, Brooks, Cole.

Guido H. E. Gendolla, et al. (2008, February). Self-focus and task difficulty effects on effort-related cardiovascular reactivity, *Psychophysiology,* Vol. 45.

Jean, G.V. (2008). Improvised explosive devices: Could they threaten U.S. ports? National Defense Magazine, Retrieved February 2, 2008, from www.nationaldefensemagazine.org/issues/2008/January/Improvised.htm.

Kozak, G. (2006). Side scan sonar target comparative techniques for port security and MCM Q-route requirements, L-3 Communications Klein Associates, Inc.

Schnedler, W. (2008). Task difficulty, performance measure characteristics and the trade off between insurance and well-allocated effort. Retrieved August 5, 2008, from www.bristol.ac.uk/cmpo/publications/papers/2006/wp147.pdf.

Truver, S.C. (2007, October). Underwater IEDs…the threat is real!, Presentation to the ASNE Flagship Seminar, Washington, D.C.

Venables, W.N., Ripley, B.D. (2002). Modern Applied Statistics With S. New York: Springer.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, VA

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, CA

3. Naval Mine and Anti-Submarine Command
   Corpus Christi, TX

4. Naval Surface Warfare Center Panama City
   Panama City, FL

5. Naval Undersea Warfare Center
   Newport, RI

6. Naval Undersea Warfare Center
   Keyport, WA